

Figure S1. Cellular composition of populations defined by flow cytometry. a, Abundance of different gates as fraction of total. **b**, Quantification of cell type composition for each FACS gate shown in main figure 1a.

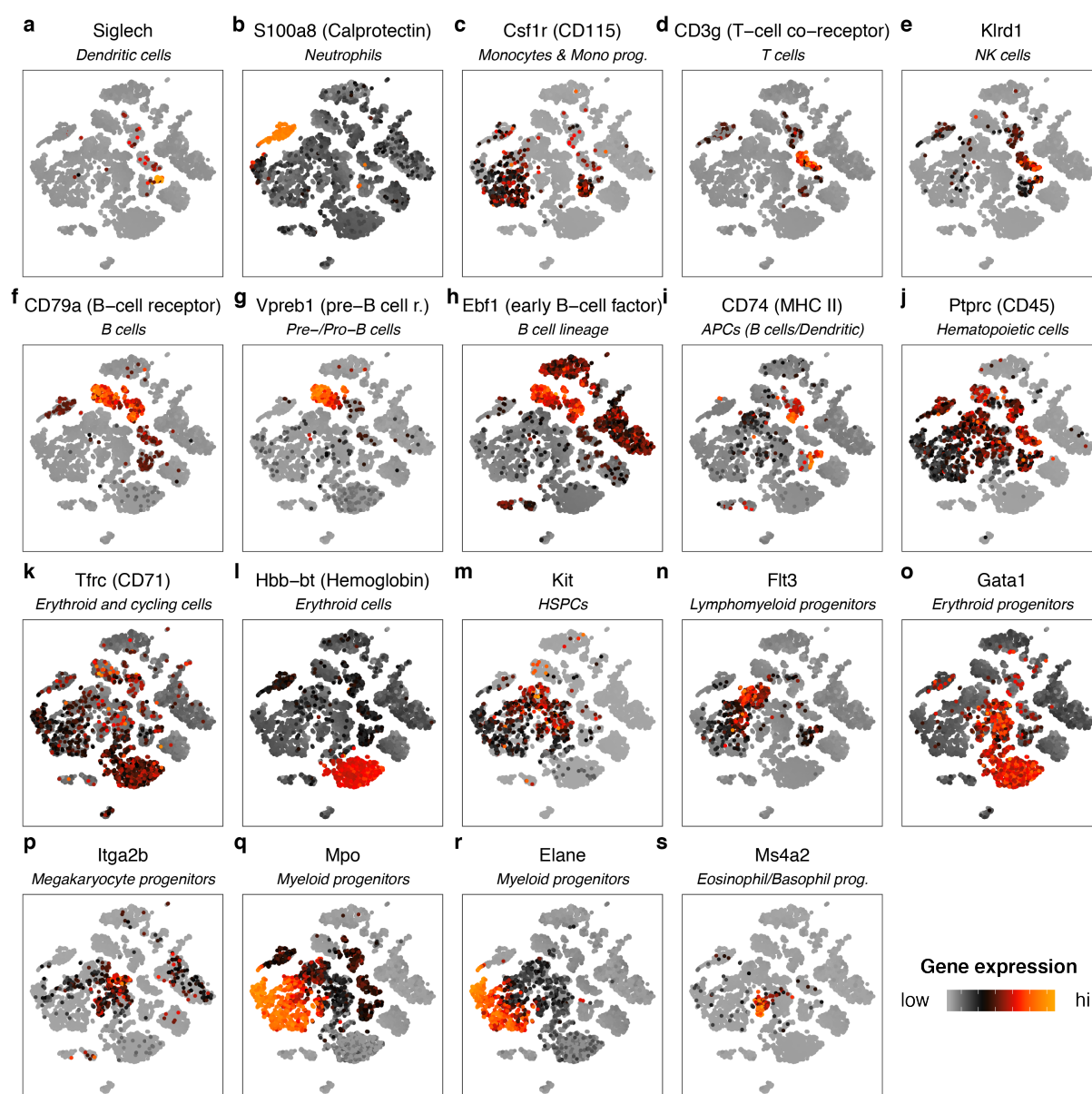


Figure S2. Expression of marker genes for haematopoietic populations highlighted on t-SNE. For full lists of marker genes, see table S1.

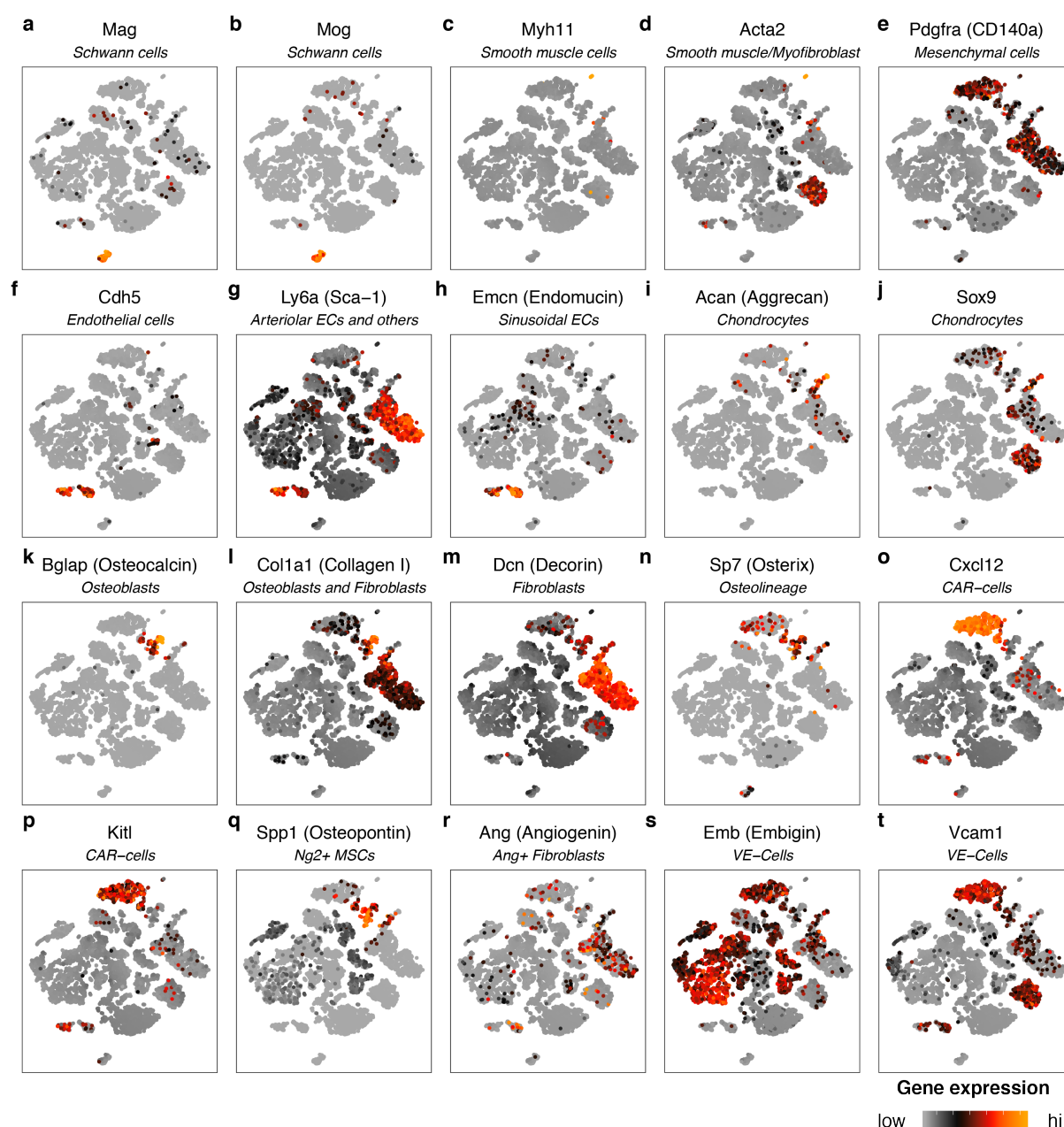


Figure S3. Expression of marker genes for non-haematopoietic populations highlighted on t-SNE. For full lists of marker genes, see table S1.

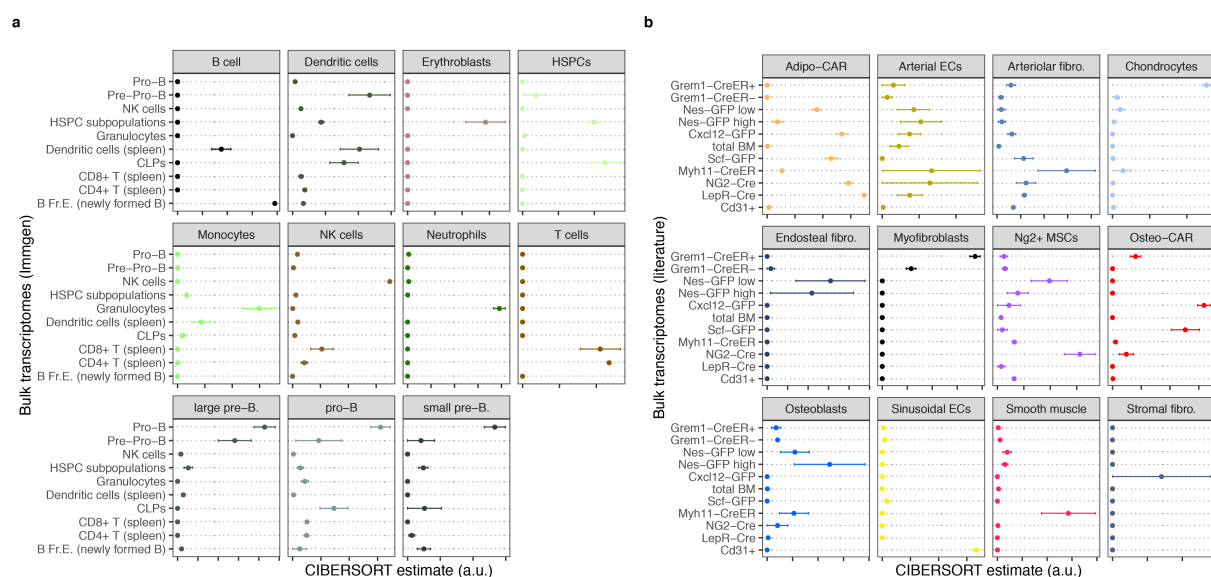


Figure S4. Comparison of cell type transcriptomes determined by scRNA-seq to data from bulk populations described in literature. **a**, Enrichment of gene expression signatures of haematopoietic populations in immune cell transcriptomes published by the immgen consortium (data source: GEO GSE109125)⁵¹. **b**, Enrichment of gene expression signatures of non-haematopoietic populations in published transcriptomes of populations defined by genetic markers^{5,7–9,52}; see methods for specification of data sources, and see the supplementary note for a detailed evaluation of the algorithm used. Error bars indicate standard error of the mean for n=3 to n=6 bulk transcriptome samples per class.

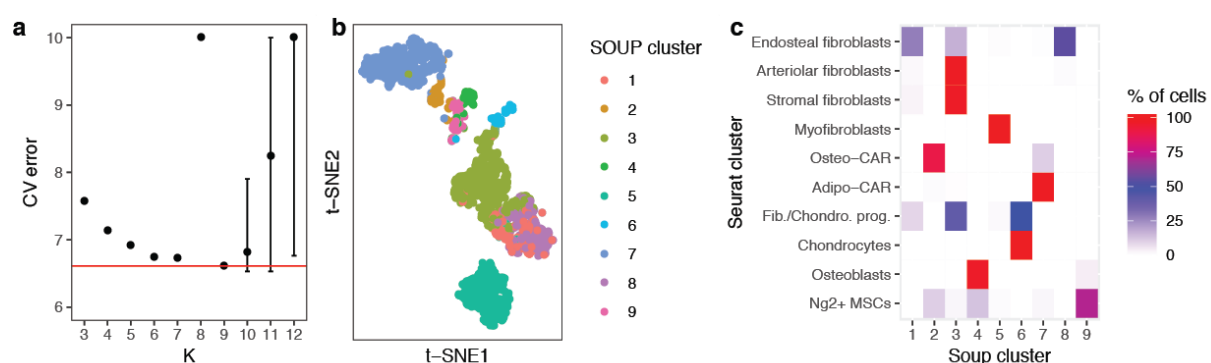


Figure S5. Comparison of clustering methods. **a**, The optimal number of mesenchymal cell clusters was determined using the SOUP method¹⁵, a semi-soft clustering algorithm designed to distinguish between distinct cell types and transition states between cell types. **b**, Main cluster identity from SOUP highlighted on the t-SNE from figure 1b (mesenchymal cell types only). **c**, Comparison of clusters identified by Seurat (Figure 1b) to clusters identified by SOUP (Figure S5b) demonstrates strong overlap between both methods.

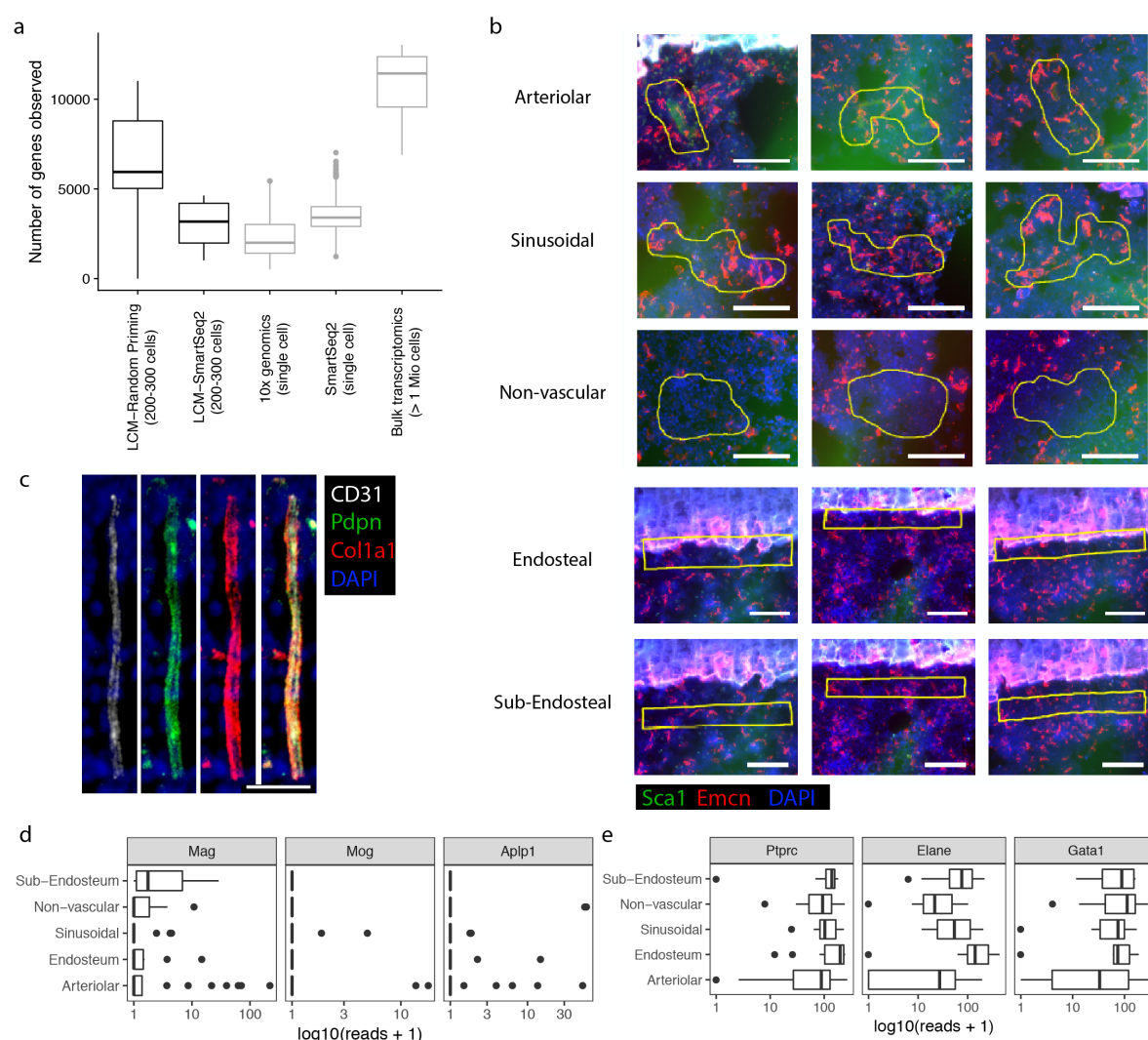


Figure S7. Technical properties of the LCM-seq dataset. **a**, Boxplots comparing the number of genes observed per sample in different protocols. All samples were down-sampled to 1 million reads for comparison. For the dataset presented in main figure 3, the protocol relying on random priming was used. **b**, Representative images of samples collected for LCM-seq; scale bar corresponds to 100 μ m. **c**, Immunofluorescence staining of a BM arteriole stained for Col1a1, Pdpn and CD31. Scale bar: 20 μ m. **d**, Schwann cell markers were lowly expressed across all niches **e**, haematopoietic markers were highly expressed across all niches.

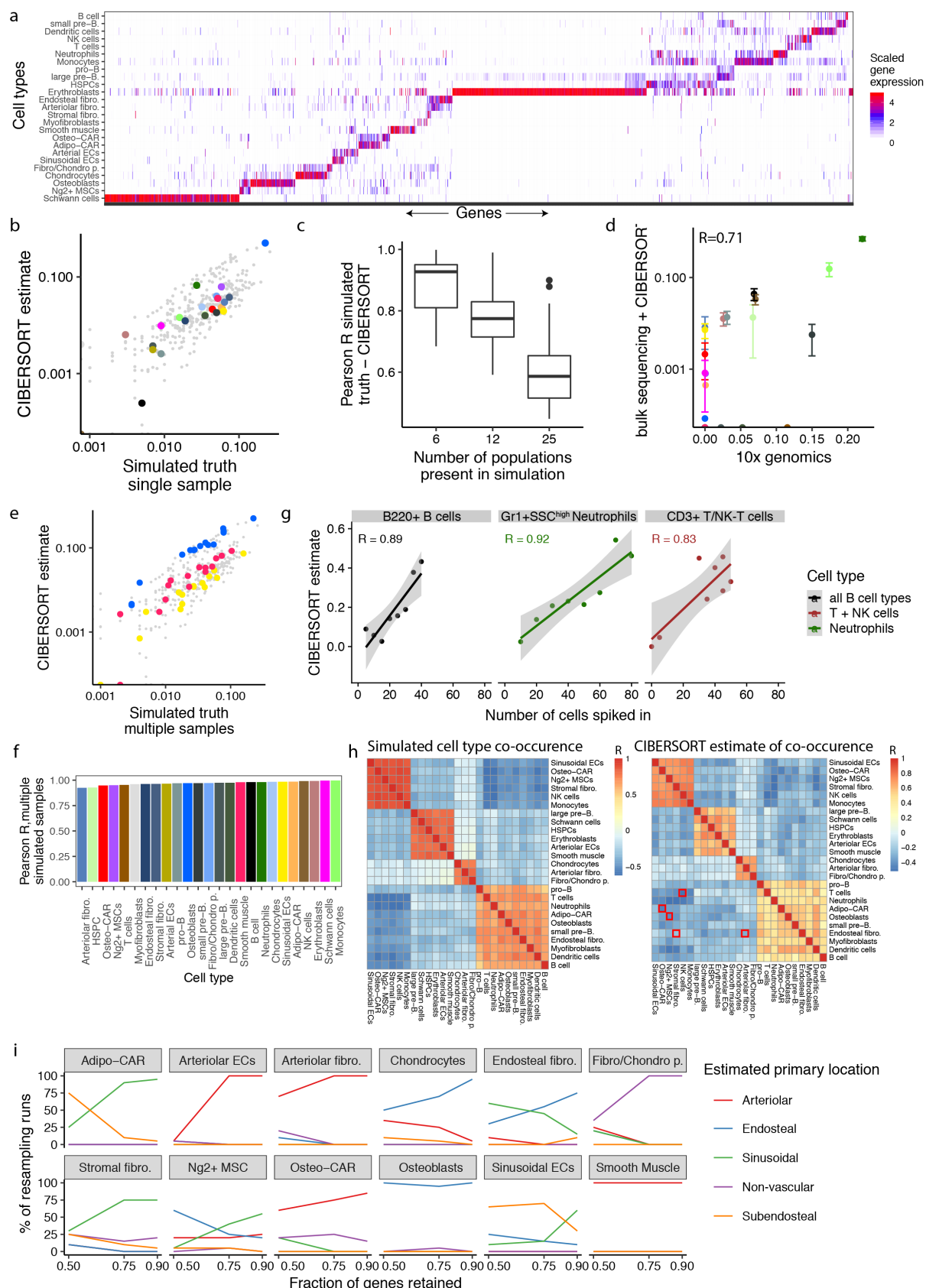


Figure S8. Evaluation of the CIBERSORT algorithm, see also supplementary note. a, Heatmap of population-specific marker genes used for the algorithm. **b,c,** Simulations to assess the ability of CIBERSORT to decompose individual samples; see supplementary note for detail.

d, CIBERSORT estimates of cell type composition of total bone marrow, compared to the cell type composition estimate from 10x genomics (see figure S1b). **e,f**, Simulations to assess the ability of CIBERSORT to identify changes in population frequencies across multiple samples; see supplementary note for detail. **g**, FACS was used to assemble 8 different pools of B220⁺ B-cells, CD3⁺ T/NK-T cells and Gr1⁺SSC^{high} neutrophils. Each pool contained a total of 100 cells at predefined ratios of B cells, T cells and neutrophils. Pools were then fixed and processed using the LCM-seq protocol, and CIBERSORT was used to decompose their composition. Estimates for T and NK cells, as well as different B-cell subpopulations, were summed for the display. **h**, Simulations to assess the ability of CIBERSORT to discriminate between similar cell types; see supplementary note for detail. Red squares highlight pairs of similar cell types. **i**, Stability of the CIBERSORT estimates from main figure 3e with regard to re-sampling of the marker gene lists used; see supplementary note for detail.

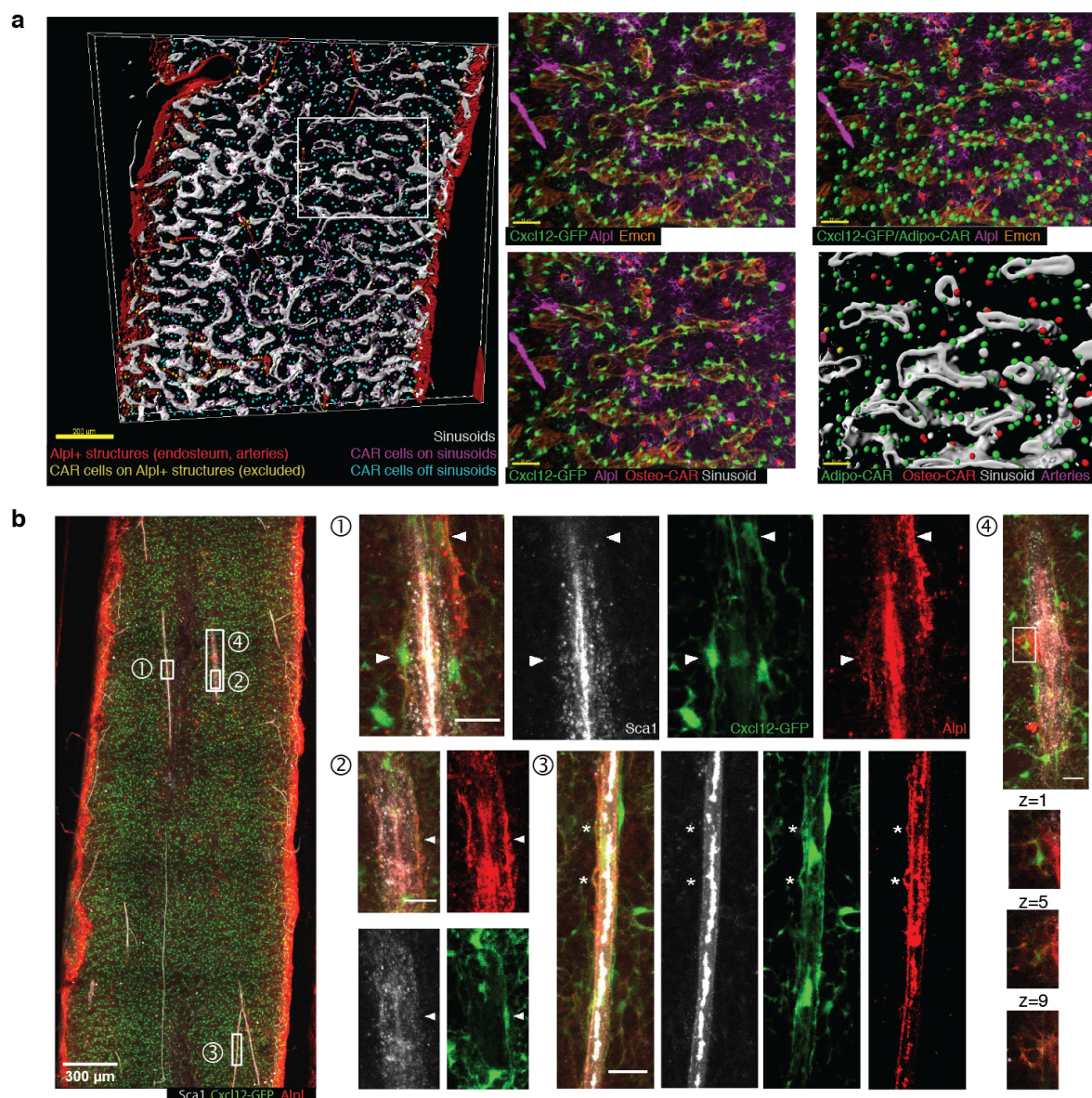


Figure S9. Whole-mount imaging and data analysis. **a**, Whole-mount imaging data of a Cxcl12-GFP bone section stained for Alpl and Emcn was segmented in 3D using the imaris software. Large Alpl⁺ surfaces (red, corresponding to endosteum and arteries) were identified and any GFP⁺ spots with <20μm proximity to these structures were excluded from further analysis (yellow spots). Remaining GFP⁺ spots were classified as within 15μm of sinusoidal vessels (purple dots), of away from sinusoidal vessels (cyan dots). GFP⁺ spots were further classified as Alpl⁺ (right panels, red spots) or Alpl⁻ (right panels, green spots). **b**, Like in main figure 4c. In ROI 3, asterisk correspond to GFP⁺Alpl⁺ protrusions on, but clearly distinct from, Sca1⁺ arteriolar endothelial cells. In ROI 4, various z-sections of a highly reticulate Cxcl12-GFP⁺Alpl⁺ cell are shown.

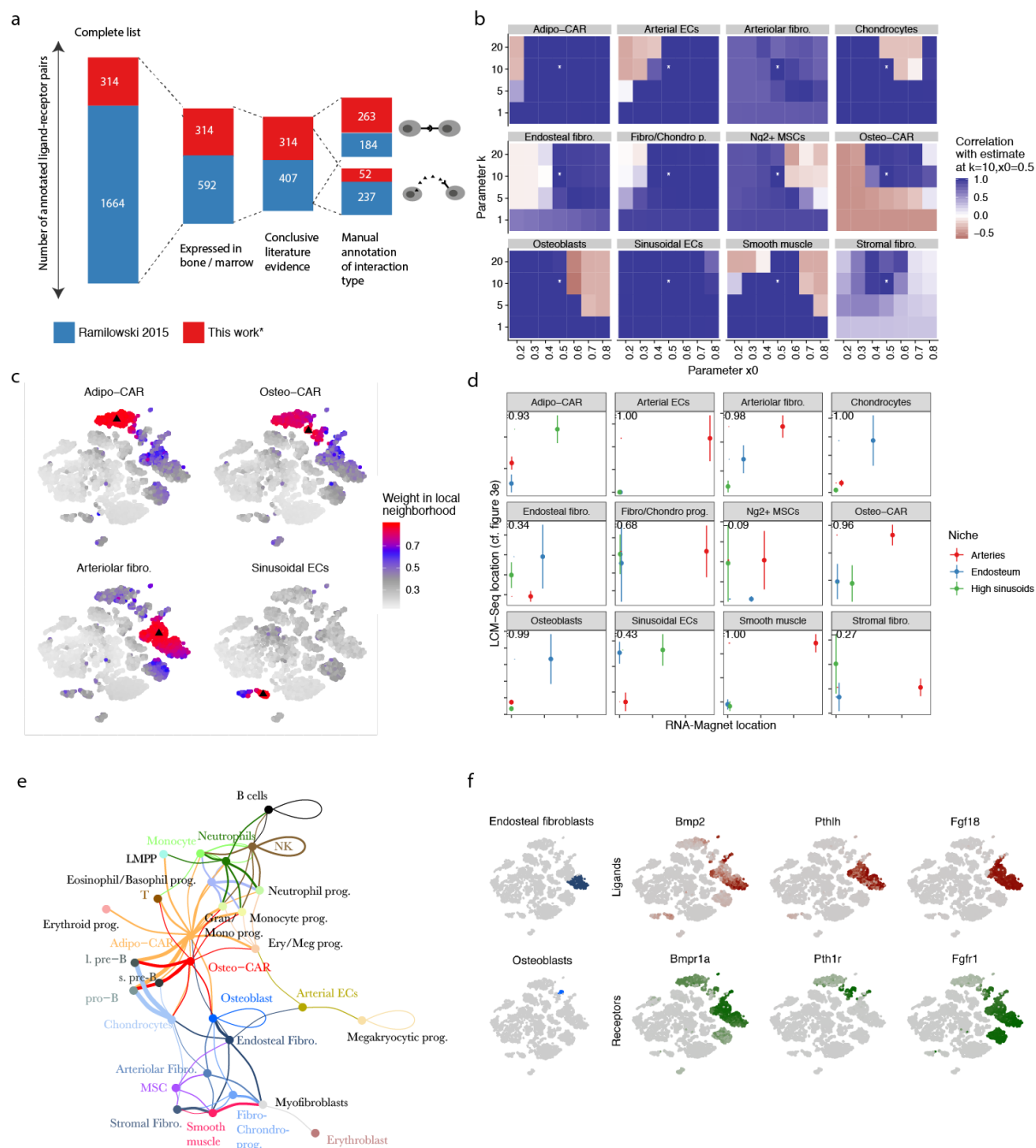


Figure S10. Analyses using RNA-Magnet. **a**, Overview of the receptor-ligand list used. See methods for data sources, and table S3 for the complete list. **b**, Stability of the RNA-Magnet location estimate for different choices of the fuzzification parameters k and x_0 . For each parameter set, RNA-Magnet location estimates were summarised per cell type, and compared to the summarised location estimate displayed in figure 5c. The asterisk indicates the parameter set used in figure 5c. **c**, Choice of local neighbourhoods. As detailed in the methods section, RNA-Magnet works by identifying interactions specific to a single cell compared to similar cells. The figure displays the size of local neighbourhoods for four representative cells

demarked by a black triangle. **d**, Detailed comparison of location estimates obtained from LCM-seq and RNA-Magnet. See also main figure 5c. **e**, Fully labelled display of the network from main figure 7a. **f**, Expression of selected cytokines and growth factors involved in bone remodelling.

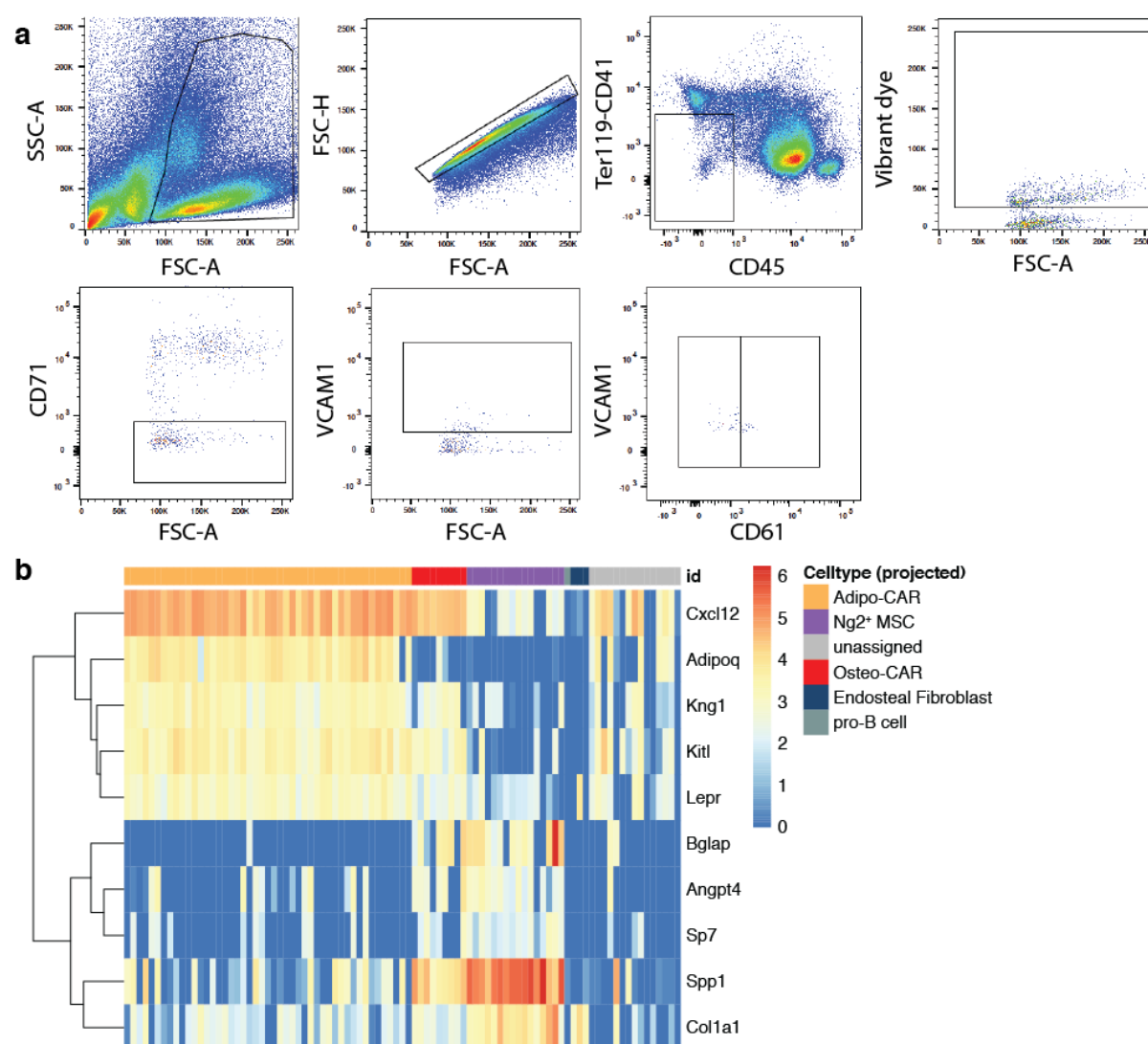


Figure S11. Index-sorting analysis of $\text{Lin}^{\text{neg}}\text{Vcam1}^+$ cells. **a**, Sorting scheme used **b**, Expression of key marker genes confirm the cell type assignment obtained by scmap, cf. main Figure 6c.

Methods

Mouse experiments

Mice were purchased from the distributors Janvier and Envigo, and housed under specific pathogen-free conditions at the central animal facility of the German Cancer Research Center. All animals used were 8-12 weeks old C56Bl/6J females. All animal experiments were

niches. We then estimated the preferred localisation of each cell c to one of these four niches N as

$$loc(c) = \underset{K \in N}{\operatorname{argmax}} S'(c, K)$$

and we estimated an ‘adhesiveness’ score for each cell based on the total number of receptors it expresses (cf. equation 1)

$$adh(c) = \sum_{r \in R} f(x_{cr})$$

For figure 7a+b, we visualized which populations specifically interact with each other by computing population-wise mean RNA-Magnet scores, and setting a threshold value above which cell types were connected in a graph.

$$R'(K, M) = \sum_{i \in M} \frac{1}{|M|} S'(i, K)$$

Finally, to obtain an estimate of total signal derived from different niches in figure 7c, we applied RNA-magnet to ligand expression data from LCM-seq.

Code availability

Our implementation of RNA-Magnet and CIBERSORT, as well as vignettes for re-creating key analysis steps are available at <http://git.embl.de/velten/rnamagnet/>

Data availability

Data are available for interactive browsing at <http://nicheview.shiny.embl.de>. Raw sequencing data and count tables are available through GEO (GSE122467, reviewer access token spqnigszdopdkh).

Supplementary Table S4. Antibodies used in this study.

Antibody	Clone	Company
Alpl	Goat Polyclonal	ThermoFisher
Anti-Goat IgG AF 546 (whole mount)	Donkey polyclonal	ThermoFisher
Anti-Rat IgG DyLight 650 (whole mount)	Donkey polyclonal	ThermoFisher
B220	RA3-6B2	eBioscience
CD105	MJ7/18	eBioscience
CD106 (VCAM1)	429	BioLegend
CD11b	M1/70	eBioscience

CD140a (PDGFRa)	APA5	eBioscience
CD144/VE-Cad	VECD1	BioLegend
CD200	OX-90	BD
CD31	390	eBioscience
CD4	RM4-4	eBioscience
CD41	eBioMWReg30	eBioscience
CD45	30-F11	eBioscience
CD51	RMV-7	eBioscience
CD61	2C9.G2	BD/BioLegend
CD71	C2	BD
CD8	53-6.7	BD
Collagen I	Rabbit polyclonal	Bio Trend
CXCL12/SDF-1	79018	R&D
Donkey Anti-Goat IgG H&L	Donkey polyclonal	Abcam
Donkey Anti-Rabbit IgG H&L	Donkey polyclonal	Abcam
Elastin	Rabbit polyclonal	Abcam
Endomucin (IF, LCM)	V.7C7	eBioscience
Endomucin (whole mount)	V.7C7	Stanta Cruz
Goat anti-rat IgG	Goat polyclonal	BioLegend
Goat Anti-Syrian hamster IgG H&L	Goat polyclonal	Abcam
Gr1	RB6-8C5	eBioscience
Podoplanin	8.1.1	BioLegend
Sca-1 (IF, LCM, FACS)	D7	eBioscience
Sca-1 (whole mount)	E13-161.7	BioLegend
SM22	Rabbit polyclonal	Abcam
Streptavidin APC-eFluor™ 780 Conjugate	-	eBioscience
Ter119	TER-119	eBioscience

1105
1106
1107

Supplementary Note

Cell type decomposition from spatial transcriptomics using CIBERSORT

CIBERSORT¹³ is an algorithm for estimating the cell type composition of a bulk sample, given a gene expression profile of the sample and a known gene expression profile for each cell type potentially contributing to the sample. Mathematically, the expected expression level x_j of gene j in a bulk sample is the sum of cell type averages, s_{ij} , weighted by cell type fractions a_i :

$$x_j = \sum_i a_i s_{ij}$$

Since the number of genes included is always much larger than the number of cell types, this formulation results in a well-determined system of linear equations. Conventional approaches for its solution however fail to distinguish similar populations and are strongly subjected to experimental noise⁷⁴. CIBERSORT avoids these problems through the use of support vector regression, which has been described to a) internally select an optimal subset of minimally correlated genes, b) penalize each cell type going into the estimate, favoring sparse solutions and c) have a linear penalty function, making it more robust against outliers driven by technical variability.

We used a per-cell type average gene expression matrix defined on 1571 genes with specificity to the individual populations (Figure S8a, genes were defined by specificity to a given population of 0.8 or greater, as quantified from areas under the ROC curve); we will discuss below how the choice of marker gene pre-selection impacts our results. To simplify analyses, we merged the highly similar HSPC subtypes into one population for CIBERSORT. In total, 25 cell types were used for all CIBERSORT analyses.

Evaluation using simulations and bulk RNA sequencing

To critically evaluate the performance of the CIBERSORT algorithm, we performed a simulation study and confirmed the results using bulk RNA-sequencing. As detailed in the following, we found that the algorithm excels at comparing relative cell type abundancies between niches (i.e. ‘cell type X localizes to niche A over niche B and niche C’), but performs only moderately at estimating cell type proportions within a single niche (i.e. it cannot draw statements like ‘niche A consists to 70% of cell type X and 30% of cell type Y’). We therefore focus our analyses to statements of the first type.

First, we evaluated the ability of CIBERSORT to estimate cell type proportions in a single niche, i.e. a single bulk RNA sequencing sample composed of the cell types described in figure 1b. For this purpose, we *in silico* created a bulk RNA sequencing sample by drawing cell type frequencies from a uniform Dirichlet distribution with 25 dimensions, resulting in a vector of cell type frequencies \mathbf{a} (ground truth). We then assumed that a pooled sample of a total of 1000 cells was to be sequenced. We sampled $1000 \times \mathbf{a}$ single cells from each population in our main dataset, and summed the gene expression values for each gene across all cells contained in the sample, resulting in a gene expression vector \mathbf{x} . This vector was then decomposed using CIBERSORT to result in an estimate of cell type frequencies $\hat{\mathbf{a}}$. We found that Pearson and Spearman correlations between the ground truth \mathbf{a} and the estimate $\hat{\mathbf{a}}$ were on the order of 0.6 (Figure S8b, c); however, populations contributing with more than 1% were identified reliably with an area under the curve (AUC) of 0.95. Correlations improved to above 0.9 if a smaller number of cell types were selected that contribute to the bulk sample, while leaving the population reference unchanged (Figure S8c).

To confirm this result, we created bulk RNA sequencing data of total bone marrow and compared the CIBERSORT estimate of its cellular composition to the estimate from our single cell RNA-seq experiment. We found that despite the different RNA-seq protocols used, the performance was as expected from our simulation study (Figure S8d; $R=0.71$, median correlation for a sample composed of 12 cell types: 0.77).

Next, we evaluated the ability of CIBERSORT to estimate changes in cell type proportions across multiple samples. For this, we repeated the sampling experiment 15 times and quantified the correlation between estimates across samples for each cell type (Figure S8e, f). An optimal performance with correlations >0.95 was found for all populations.

To confirm this result, we used FACS to assemble 8 different pools of B220+ B-cells, CD3+ T/NK-T cells and Gr1+SSC^{high} neutrophils. Each pool contained between 5 and 80 cells of each type, for a total of 100 cells. Pools were then fixed and processed using the same protocol used for the laser microdissected samples, and CIBERSORT was used to quantify their composition. As expected from the simulation study, changes in cell type proportions across samples were very accurately identified with a Pearson R of 0.83-0.92 (Figure S8g).

In line with previous studies^{13,75}, these analyses suggest that CIBERSORT excels at identifying changes in cell type proportion across multiple samples but performs only moderately at estimating cell type proportions in a single sample. We therefore restrict the use of CIBERSORT to comparing relative cell type abundancies between niches, and do not determine absolute cell frequencies.

Finally, we also evaluated the extent to which CIBERSORT is capable of discriminating cell types that exhibit similar gene expression profiles (e.g. distinct fibroblast populations or Adipo- and Osteo-CAR cells). We therefore simulated 100 bulk samples assuming that cell types co-occur in a pre-specified manner. Mathematically, we first manually specified a correlation structure of cell type co-occurrence \mathbf{C} (of dimensions 25x25). We then sampled a cell type frequency matrix \mathbf{A} (of dimensions 100x25) with a covariance structure \mathbf{C} . Importantly, correlations were thereby specified and sampled at the level of cell types, and not at the level of genes. We then created a bulk RNA expression profile *in silico* for each sample (row in \mathbf{A}) as described above, and applied CIBERSORT to estimate its cellular composition. This resulted in a matrix of estimated cell type proportions $\hat{\mathbf{A}}$. Column correlation structures of \mathbf{A} and $\hat{\mathbf{A}}$ are compared in figure S8h. Importantly, cell type co-occurrence is correctly identified and not influenced by similarity in the gene expression profile of the reference populations.

In summary, the simulation and bulk RNA sequencing study performed here supports previous evaluations of CIBERSORT: The algorithm is ideally suited for identifying changes in cellular composition between multiple samples. A more detailed analyses of its performance e.g. with regard to noise can be found in ref. ¹³.

Impact of marker gene selection on CIBERSORT results

While CIBERSORT internally selects an optimal set of marker genes, it also requires the pre-specification of a set of reasonably specific markers. To gauge the dependence of CIBERSORT results on marker genes, we repeated the analyses of LCM samples (Figure 2) 60 times, each time using a random subset of 50%, 75% or 90% of the marker genes of each population. For each cell type, we subsequently quantified the fraction of resampling runs that result in the same primary location. The result (Figure S8i) allows us to assess the stability of the CIBERSORT estimates as follows:

For Osteoblasts (n=108 marker genes) and Smooth muscle cells (n=82), any 50% of marker genes can be left out while still allowing unanimous placement of these cells at the endosteum or arteries, respectively.

For Arteriolar fibroblasts (n=54), Arteriolar endothelial cells (n=50) and Fibroblast-Chondrocyte precursors (n=37), any 25% of marker genes can be left out while still allowing unanimous placement of these cells at their respective locations.

For Adipo-CAR cells (n=71), Osteo-CAR cells (n=61), Chondrocytes (n=86) and stromal fibroblasts (n=26), if 25% of marker genes are dropped, this resulted in location swaps in between 10% and 30% of cases. However, the swap was mostly between the primary and a potential secondary location of the cells.

For Sinusoidal endothelial cells (n=33), endosteal fibroblasts (n=69), and MSCs (n=23), estimates depended more strictly on lists of marker genes used. Small numbers of specific markers, elevated intra-population heterogeneity and/or a more ubiquitous localization of these cells may be factors contributing to the estimation uncertainty. For sinusoids and endosteal fibroblasts, we provide further evidence for their localization in figures 2b and 4e, respectively.